Vol. 59 No. 6 June 1984

JOURNAL OF Medical Education

PUBLICATION OF THE ASSOCIATION OF AMERICAN MEDICAL COLLEGES

MCAT and the Selection of Medical Students

Validity of the MCAT in Predicting Performance in the First Two Years of Medical School
Robert F. Jones and Maria Thomae-Forgues

The Future of Standardized Testing (Editorial)
James B. Erdmann

MCAT Scores and Student Progress in Medical School (Datagram)
Robert F. Jones and Suzanne Vanyur

Development of Clinician-Educator Faculty Track at the University of Pennsylvania
Maureen Parris and Edward J. Stemmler

The Quality of Deans' Letters from Medical Schools
Joel Yager, Gordon D. Strauss, and Kenneth Tardiff

Educational Influences on Consultation Rates of House Staff Physicians in a Primary Care Clinic
George D. Everett, Thomas J. Parsons, and Andrea L. Christensen

Journal of Medical Education

Volume 59 • Number 6 • June 1984

Validity of the MCAT in Predicting Performance in the First Two Years of Medical School—Robert F. Jones and Maria Thomae-Forgues 455

Development of Clinician-Educator Faculty Track at the University of Pennsylvania—Maureen Parris and Edward J. Stemmler 465

The Quality of Deans' Letters from Medical Schools—Joel Yager, Gordon D. Strauss, and Kenneth Tardiff 471

Educational Influences on Consultation Rates of House Staff Physicians in a Primary Care Clinic—George D. Everett, Thomas J. Parsons, and Andrea L. Christensen 479

Measuring the Complexity of Clinical Problems—Brian K. Hennen 487

An Agenda for Residency Training in Ambulatory Care—Kurt Link and David Buchsbaum 494

Knowledge of Medical Students, Residents, and Attending Physicians About Opiate Abuse—Daniel Shine and Penelope Demas 501

Reducing 'Information Overload' in the Teaching of Pharmacology: The '200 Drug List'—Merle W. Riley 508

Communications: A Data Base for Curriculum Design in Medical Ethics—Richard G. Tiberius and Doreen Cleave-Hogg 512

Recognition of Outstanding Performance in Medical Schools with Nontraditional Grading Systems—M. J. Peters and W. Tyree Finch 514

Effect of Attendance at Lectures on Medical Student Performance—Karen Glanz and Stanley B. Fiel 516

Introducing Dietary Assessment to First-Year Medical Students—Susie Bredderman, Linda Nieman, and Ann Cox 518

Microcomputers for Computer-Assisted Instruction in Anesthesiology—H. Steele Holley and Floyd N. Heller 521

Briefs 522

Editorial: The Future of Standardized Testing 525

Datagram: MCAT Scores and Student Progress in Medical School 527

Bibliography 532

Letters to the Editor 536

Officers of the Association and Members of the Executive Council, 1983–84

CHAIRMAN: Robert M. Heyssel, *Johns Hopkins Hospital*
CHAIRMAN-ELECT: Richard Janeway, *Bowman Gray*
IMMEDIATE PAST CHAIRMAN: Steven C. Beering, *Purdue University*
PRESIDENT: John A. D. Cooper, *AAMC, Washington, D.C.*

COUNCIL OF DEANS
Edward J. Stemmler, *University of Pennsylvania**
Arnold L. Brown, *University of Wisconsin*
John E. Chapman, *Vanderbilt University*
D. Kay Clawson, *University of Kansas*
Fairfield Goodale, *Medical College of Georgia*
Louis J. Kettel, *University of Arizona*
Richard H. Moy, *Southern Illinois*
John Naughton, *SUNY, Buffalo*

COUNCIL OF ACADEMIC SOCIETIES
Robert L. Hill, *Duke University**
Joseph E. Johnson, III, *Bowman Gray*
Virginia V. Weldon, *Washington University*
Frank C. Wilson, *University of North Carolina*

COUNCIL OF TEACHING HOSPITALS
Haynes Rice, *Howard University Hospital**
Robert Frank, *Barnes Hospital*
Earl J. Frederick, *Children's Memorial Hospital*
Sheldon S. King, *Stanford University Hospital*

ORGANIZATION OF STUDENT REPRESENTATIVES
Pamelyn Close, *University of Tennessee**
Ricardo Sanchez, *Brown University*

DISTINGUISHED SERVICE MEMBER
Charles C. Sprague, *University of Texas, Dallas*

* Chairman of respective organization.

Validity of the MCAT in Predicting Performance In the First Two Years of Medical School

Robert F. Jones, Ph.D., and Maria Thomae-Forgues

Abstract—In this paper, the authors present the first systematic summary of predictive validity research on the new Medical College Admission Test (MCAT) since its introduction in 1977. Data are drawn primarily from the MCAT Interpretive Studies Program, a cooperative effort between the Association of American Medical Colleges and 30 of its member schools to conduct research that will both facilitate local use of the test scores and contribute to a national perspective on their value in medical school admissions. The results show that MCAT scores by themselves have significant predictive validity with respect to first- and second-year medical school course grades and National Board of Medical Examiners Part I examination scores and that they complement the predictive validity of undergraduate college grades. The MCAT Science Knowledge areas of assessment, particularly Biology and Chemistry, and the Science Problems subtest tend to have higher correlations than the Skills Analysis subtests with initial performance in medical school; however, the Skills Analysis: Reading subtest may retain its predictive value best over time. Correlation values are discussed in terms of methodological factors which constrain their size. They are also compared with those found for other professional and graduate school admission tests. Further directions for MCAT validity research are described.

Beginning with the first use of the new Medical College Admission Test (MCAT) for selecting entrants to medical school in 1978, there has been great interest among admissions officers and researchers in studies designed to assess the test's predictive validity. Eleven such studies appeared in the *Journal of Medical Education* alone during 1979–1983, while many more were the topic of presentations at professional meetings (1). The Association of American Medical Colleges (AAMC), sponsor of the MCAT, has shared this interest. In 1980 the AAMC began the MCAT Interpretive Studies Program, a cooperative effort with 30 selected schools of medicine to conduct validity research on the MCAT. The purpose of the authors in this paper is to provide a preliminary summary of the MCAT validity evidence as it relates to performance in the first two years of medical school.

There are as many questions that can be addressed concerning MCAT validity

Dr. Jones is a research associate in the Division of Educational Measurement and Research of the Association of American Medical Colleges (AAMC), Washington, D.C. Ms. Thomae-Forgues is a human resources specialist at the Organization of American States, Washington, D.C.; she formerly was a research associate in the Division of Student Programs at the AAMC.

For related editorial and datagram, see pages 525 and 527.

as there are specific inferences to be made from the scores. The authors' purpose in this summary is to address some validity questions of general interest. The answers may not apply to each situation but reveal general patterns in the ability of MCAT scores to predict the performance of students in the basic medical sciences. Five commonly posed questions or sets of questions in this regard are:

1. How does the predictive validity of MCAT scores in relation to performance in the basic medical sciences compare with that of the undergraduate grade-point average (GPA)?
2. Do the MCAT scores contain unique information—not already provided by undergraduate GPA—that aids in the prediction of students' performance? To what degree?
3. What is the relative predictive validity of the individual MCAT scores in relation to overall performance in the basic medical sciences?
4. What is the relative predictive validity of the individual MCAT scores in relation to performance in specific areas?
5. How strong is the predictive validity of MCAT scores? Does it meet the standards that should be expected of a measure of its kind? How does it compare with that of other graduate and professional school admission tests?

Procedures

To answer these questions, results of studies were available from half of the 30 schools participating in the MCAT Interpretive Studies Program. To increase the sample, results from studies which were comparable in terms of the statistical indices used and which were from five schools not participating in the MCAT Interpretive Studies Program were added to the data from the 15 schools. Eighteen of the 20 schools are United States medical schools, 13 public and five private; two are Canadian schools. Sixteen of the schools have a traditional, discipline-oriented curriculum in the first two years, two use an organ systems approach, and two provide a mix of straight discipline courses and organ systems units. The studies were conducted separately by class, primarily those entering in 1978 and 1979. Most of the schools contributed data from more than one class, but not all types of criterion measures or statistical indices were available from all schools.

Criterion performance measures for the first two years were classified along several lines. Summary measures of first-year or second-year performance were designated year 1 grades and year 2 grades, respectively. These included criteria such as grade-point average for courses taken in a given year, class rank, or scores on an end-of-year comprehensive examination. These criteria are important for MCAT validity studies because they reflect locally determined standards of student performance. Total scores on the National Board of Medical Examiners (NBME) Part I examination provide another summary measure of performance in the basic medical sciences. The NBME Part I examination derives its importance as a criterion for MCAT validity studies from its current role in procedures for physician licensure. Subtests of the NBME Part I examination were also used as criteria to study relationships between MCAT scores and performance in specific disciplines. Local course grades or examination scores were grouped similarly by the disciplines represented in the NBME Part I examination subtests. Grades in interdisciplinary courses were too infrequently used as criteria to be examined separately in this summary.

Two statistics commonly used in predictive validity studies to index the relationship between test scores and criterion measures provided the raw data for this

preliminary summary. The first was the simple Pearson product-moment correlation coefficient (r) or validity coefficient. Pearson correlations were computed for each MCAT scale and each undergraduate GPA measure (science, nonscience, total) with each criterion measure whose underlying measurement scale was more than simply categorical (for example, honors-pass-fail grades were not analyzed in this fashion).

Since these correlations were based on enrolled student samples which, by virtue of the selection process, are quite homogeneous with respect to the predictive measures being studied, they were then corrected for restriction of range by a commonly used formula (2). Corrected correlations attempt to estimate what the observed correlations would be if all applicants to the school, not just the more homogeneous sample of enrolled students, were studied.

The second index used was the multiple correlation coefficient. This is the Pearson correlation that results when a group of predictor variables are combined in an optimal way and related to a criterion measure. Multiple correlations were computed with all summary criterion measures (year 1 grades, year 2 grades, and NBME Part I examination score total) for (*a*) the six MCAT scores combined, (*b*) the two independent GPA measures (science and nonscience) combined, and (*c*) the MCAT scores and GPA measures combined. The size of these multiple correlations is similarly constrained by the limited range of performance on the predictive measures shown in the enrolled student samples. However, a correction formula for multiple correlations is not available.

Results and Discussion

Table 1 displays the distributions of multiple correlations computed with year 1 grades, year 2 grades, and NBME Part I examination scores, while Table 2 compares the percentage of times a GPA or MCAT composite was the better predictor of performance. Table 3 displays the average individual correlations or validity coefficients for these same criteria. Table 4 gives the percentage of times individual GPA and MCAT scores were the best single predictors of performance. Table 5 shows the average validity coefficients for criteria classified by basic science area and method of evaluation. Even at this early stage in the accumulation of data, certain patterns or trends are evident which relate to the questions posed previously.

Tables 1 and 2 show that whether MCAT scores or undergraduate GPAs were better predictors of performance in the first two years of medical school depended upon the criterion being considered. When the criteria were medical school course grades, MCAT scores in combination were similar to undergraduate grades in their predictive value. Median multiple correlations were identical, and within any one sample either predictor group was about equally likely to show a higher multiple correlation. However, as shown in Table 3, no single MCAT score tended to be correlated with medical school course grades as highly as undergraduate science GPA. When the criteria were NBME Part I examination scores, MCAT scores in combination (Table 1) were substantially better predictors of performance than undergraduate grades. This was true of every sample studied (Table 2). Moreover, several individual MCAT subtest scores (Table 3) were the best single predictors of NBME examination performance in terms of their average correlation.

The data in tables 1 and 2 also bear on the second validity question which concerns the degree to which MCAT scores contribute unique and useful information

TABLE 1

Distributions of Multiple Correlations for Grade-Point Average (GPA), Medical College Admission Test (MCAT), and GPA/MCAT Composites with Year 1 Grades, Year 2 Grades, and NBME Part I Examination Scores*

Predictor Composite	Grades		NBME Part I
	Year 1	Year 2	
GPA†			
Third quartile	.46	.42	.37
Median	.41	.37	.30
First quartile	.34	.27	.23
MCAT			
Third quartile	.49	.44	.63
Median	.41	.37	.54
First quartile	.34	.29	.43
GPA/MCAT			
Third quartile	.58	.56	.68
Median	.52	.51	.59
First quartile	.47	.40	.48

* Year 1 grades are based on 25 samples (classes) at 12 schools; year 2 grades, 22 samples at 12 schools; and NBME Part I examination scores, 18 samples at 9 schools.

† Science GPA and nonscience GPA.

TABLE 2

Number and Percentage of Samples (Classes) for Which GPA, MCAT, and GPA/MCAT Composites Were Better Predictors of Year 1 and Year 2 Grades and NBME Part I Examination Scores*

Predictor Composite	Grades					
	Year 1		Year 2		NBME Part I	
	No.	Percent	No.	Percent	No.	Percent
GPA better than MCAT	13	52	10	45	0	0
MCAT better than GPA	12	48	12	55	18	100
GPA/MCAT better than GPA alone	25	100	22	100	18	100

*Year 1 grades are based on 25 samples (classes) at 12 schools; year 2 grades, 22 samples at 12 schools; and NBME Part I examination scores, 18 samples at 9 schools.

to the admissions process. In theory, MCAT scores and undergraduate grades are complementary pieces of information. The shortcomings of one measure are the strengths of the other. The MCAT allows the assessment of students on a standard content of knowledge and skills and the reporting of scores on a standard scale but is limited to sampling performance on a single day in a somewhat artificial setting. GPA is not standardized in either way but is based on repeated assessments of a student's performance over a period of time. In the present study, the multiple correlation values provided empirical support for this complementary relationship. Those based on a combination of MCAT scores and undergraduate GPA were consistently higher than those based on either predictor group separately. The increase in the average multiple correlation when the MCAT scores were added to the GPA was .11 to .14 when course grades were

the criteria and .29 when NBME examination scores were the criteria. These comparisons are usually expressed in terms of the "proportion of variance explained," that is, the multiple correlation values squared. In these terms, the MCAT scores improved predictability by as much as 90 percent with course grades as the criteria and by nearly 300 percent with NBME examination scores. Moreover, the contribution of the MCAT to predictability generalized across samples. In all cases, the increase in the multiple correlation value when the MCAT scores were added to the GPA was statistically significant.

The third question concerning how individual MCAT scores relate to overall performance in the basic sciences is addressed by the correlation values, observed and corrected for range restriction, shown in Table 3. Higher average correlations with year 1 grades were found for the MCAT Science Knowledge and Science Problems subtests than for the Skills Analysis subtests. Chemistry had the highest average correlation, with the Biology assessment area and the Science Problems subtest being only slightly less. In over two-thirds of the samples studied, either the assessment area of Chemistry or of Biology was the best predictor among MCAT scores (Table 4). Physics was the one exception to the predominantly higher correlations found for the Science Knowledge areas of assessment than the Skills Analysis areas of assessment. Its average correlation with year 1 grades was lower than the other science assessment areas, and it correlated best among MCAT scores with year 1 grades in only one of the 34 samples (3 percent).

A different pattern of results is shown with year 2 grades as the criterion. Correlations between MCAT scores and year 2 grades were all systematically lower, except for the Skills Analysis: Reading subtest. As a result, the average correlation for the Skills Analysis: Reading subtest with year 2 grades was on the same level as the correlations for the Biology and Chemistry assessment areas and the Science Problems subtest. Moreover, in 25 percent of the samples studied, the

TABLE 3

Median Individual Correlations, Observed and Corrected for Range Restriction, for Undergraduate GPA and MCAT Scores with Year 1 Grades, Year 2 Grades, and NBME Part I Examination Scores*

Predictor	Grades					
	Year 1		Year 2		NBME Part I	
	Observed	Corrected	Observed	Corrected	Observed	Corrected
Undergraduate GPA						
Science	.36	.54	.35	.47	.29	.43
Nonscience	.21	.26	.22	.26	.14	.17
Total	.37	.50	.34	.45	.25	.35
MCAT						
Biology	.34	.38	.23	.29	.40	.50
Chemistry	.31	.42	.23	.28	.43	.56
Physics	.26	.29	.14	.16	.34	.37
Science Problems	.31	.39	.22	.27	.43	.50
Skills Analysis: Reading	.19	.26	.21	.28	.24	.32
Skills Analysis: Quantitative	.24	.27	.16	.21	.29	.38

* Year 1 grades are based on 34 samples at 18 schools; year 2 grades, 28 samples at 16 schools; and NBME Part I examination scores, 26 samples at 14 schools. Larger numbers than in Table 1 reflect the availability of data from published studies.

TABLE 4

Number and Percentage of Samples (Classes) for Which Individual GPA and MCAT Scores Were Best Single Predictors of Year 1 Grades, Year 2 Grades, and NBME Part I Examination Scores*

Predictor	Grades					
	Year 1		Year 2		NBME Part I	
	No.	Percent	No.	Percent	No.	Percent
Undergraduate GPA						
Science	28	82	18	64	21	81
Nonscience	0	0	0	0	0	0
Total	6	18	10	36	5	19
Total samples	34	100	28	100	26	100
MCAT						
Biology	11	32	6	21	5	19
Chemistry	12	35	8	29	13	50
Physics	1	3	1	4	0	0
Science Problems	5	15	2	7	7	27
Skills Analysis: Reading	3	9	7	25	1	4
Skills Analysis: Quantitative	2	6	4	14	0	0
Total samples	34	100	28	100	26	100

* Year 1 grades are based on 34 samples at 18 schools; year 2 grades, 28 samples at 16 schools; and NBME Part I examination scores, 26 samples at 14 schools.

Skills Analysis: Reading subtest was the best MCAT predictor of year 2 grades, only slightly less than the percentage for the Chemistry assessment area. These findings appear to reflect the persistent nature of the skills differences shown in the Skills Analysis: Reading scores. Differences in Science Knowledge scores that students exhibit on the MCAT are reduced by the time they complete the first year of the basic science curriculum. These differences are then less useful for predicting the relative performance of students in the second year. The medical school curriculum presumably has a less direct impact in reducing differences in the more basic skills measured by the Skills Analysis: Reading subtest. Therefore, while the Science Knowledge and Science Problems subtests are better predictors of how students perform initially in medical school, the Skills Analysis: Reading subtest may be one that is more enduring.

The Science Knowledge and Science Problems subtests showed the highest correlations with performance on NBME Part I examination scores. In half of the samples studied, the Chemistry assessment area was the best MCAT predictor. In the remaining samples, usually either the Science Problems subtest or the Biology assessment area was the best MCAT predictor.

Data in Table 5 address the fourth question of how MCAT scores relate to specific areas of the curriculum. First, the data reinforce a point made earlier in this paper. MCAT correlations within each basic science area were consistently higher when performance was measured by subtests of the NBME Part I examination. This occurred despite the fact that anatomy, physiology, and biochemistry courses are taken typically in the first year of medical school, a full year prior to sitting for Part I of the NBME examinations. Undergraduate GPA tended to be more highly correlated in each area with local course grades than with NBME examination scores.

MCAT Science Knowledge subtest and

TABLE 5
Median Individual Correlations (Corrected for Range Restriction) for Undergraduate GPA and MCAT Scores with Grades and NBME Scores in Specific Disciplines

Predictor	Anatomy		Physiology		Biochemistry		Pathology		Microbiology		Pharmacology		Behavioral Science	
	Grade	NBME	Grade	NBME	Grade	NBME	Grade	NBME	Grade	NBME	Grade	NBME	Grade	NBME
Undergraduate GPA														
Science	.40	.35	.47	.38	.44	.36	.42	.36	.43	.34	.40	.40	.31	.15
Nonscience	.21	.10	.24	.08	.25	.08	.23	.20	.24	.10	.22	.18	.26	.16
Total	.39	.29	.42	.29	.39	.29	.41	.35	.39	.27	.40	.29	.34	.18
MCAT														
Biology	.27	.40	.35	.47	.31	.39	.24	.35	.29	.38	.19	.30	.21	.38
Chemistry	.21	.43	.45	.61	.41	.54	.25	.37	.23	.40	.28	.44	.19	.37
Physics	.17	.26	.37	.48	.23	.32	.10	.16	.08	.23	.17	.25	.12	.34
Science Problems	.21	.40	.45	.59	.36	.46	.21	.35	.25	.38	.23	.39	.21	.39
Skills Analysis: Reading	.11	.17	.22	.27	.14	.16	.21	.25	.21	.26	.23	.19	.34	.51
Skills Analysis: Quantitative	.14	.26	.33	.42	.23	.27	.12	.18	.19	.21	.15	.22	.21	.44

Science Problems subtest scores showed particularly strong correlations with performance in two areas of the curriculum: physiology and biochemistry. In fact, the one exception to the generally lower correlations for the MCAT Physics assessment area compared with other Science Knowledge assessment areas was in terms of performance in physiology. The pattern of correlations between MCAT scores and performance in specific areas tended to be consistent with content similarities. The MCAT Biology assessment area had higher correlations than other MCAT assessment areas with course grade performance in anatomy and microbiology. The MCAT assessment area of Chemistry had higher correlations with performance in biochemistry and pharmacology. In addition to the MCAT assessment area of Chemistry and the Science Problems subtest, MCAT Biology and Physics assessment areas and the Skills Analysis: Quantitative subtest each showed high correlations with performance in physiology. Finally, the Skills Analysis: Reading subtest was the best predictor of performance in behavioral science courses and the Behavioral Sciences subtest of the NBME Part I examination.

These comparisons of correlation values show the relative predictive validity of the MCAT subtests in relation to several different performance criteria. Of further interest is the magnitude of these correlations and what this implies about the strength of predictive validity. Any predictive measure with a validity coefficient greater than zero provides some advantage over a random process in selecting students who will perform well. But is the size of these correlations what one would hope for in a measure of this kind?

This question can be addressed only after recognizing several factors in these

studies which artificially constrain the size of the validity coefficients obtained. These are (*a*) effects due to selection which render the enrolled student samples unrepresentative of the corresponding applicant pools in terms of range of ability; (*b*) effects of limited variance on the criterion measure; and (*c*) effects of low reliability of the criterion measure. The authors partially dealt with the first factor, referred to as the restriction of range problem, by correcting the individual validity coefficients. However, recent research (3) indicates that even these upward adjustments are probably too conservative. They do not provide a substitute for the ideal experiment of studying students randomly selected from the applicant pool. Golmon and Berry (4) in their study of students at Northwestern University Medical School have come closest to this ideal situation and have shown the dramatic differences that result.

The second factor, the effects of limited variance on the criterion measure, also was partially dealt with in this study by performing correlational analyses only for those criterion measures with a minimally adequate distribution of student performance. However, Sedlacek and Hutchins (5) demonstrated that even rather small differences between samples with regard to criterion variance are reflected in the size of the validity coefficients obtained.

The third factor, the potentially low reliability of certain criterion measures, is based on the psychometric tenet that validity cannot exceed reliability. This applies particularly to course grades or local examination scores, whose reliability was not known to the authors. While many of the criteria used in these studies may have high reliability, the presence of even some with low reliability would serve to lower the average of the coefficients obtained.

With these considerations in mind, the authors can describe certain levels of strength in the predictive validity of MCAT scores. Multiple and individual correlations with NBME examination scores reflect extremely strong predictive validity. In general, these indices rarely exceed the values shown for measures of this kind. Correlations with first-year grades demonstrated fairly strong predictive validity, particularly with regard to performance in physiology and biochemistry courses. The incremental predictive validity described earlier of MCAT scores combined with GPA was the primary consideration here. The MCAT/GPA combination achieved a level of predictive validity with medical school course grades only slightly less than with NBME Part I examination scores. Correlations between individual MCAT subtests and second-year grades reflected only moderate to weak predictive validity. However, the predictive validity of the MCAT score composite tended to be in the moderate to fairly strong range, which might encompass coefficients of .3 to .5.

MCAT validity results tend to be comparable with those found for the other graduate and professional school admission test programs. For example, in one study (6), the average observed validity coefficients for the Law School Admission Test (LSAT), with first-year law school grades as the criterion, tended to be in the .3 to .4 range. Correlations between scores on the Graduate Record Examination (GRE) Aptitude Tests and graduate school grades varied by area of study but averaged approximately .3. Those for the GRE Advanced Tests tended to be higher but were still only in the .3 to .4 range. Results for the Graduate Management Admission Test were similar, with median validity coefficients being approximately .3, according to a report by Lawrence W. Hecht and Donald E. Powers, "The Predictive Validity

of Preadmission Measures in Graduate Management Education: Three Years of the GMAC Validity Study Service," presented as a symposium paper at the annual meeting of the American Educational Research Association, New York City, March 1982. Each of these tests was validated under similar constraints as the MCAT, and the results need to be viewed accordingly.

Conclusion

The MCAT Interpretive Studies Program is a major effort by the AAMC to conduct research on the test used in medical school admissions. The results of validity studies conducted through the program thus far tend to support the MCAT's value as a predictive measure in the medical school admission process and to support its continued use to the degree that the criterion performance measures that have been studied are deemed important. A systematic examination of research results has revealed some obvious trends. The authors expect that the identification of these trends will give perspective to validity research being conducted now and in the future. The exceptions to these trends are no less important. As more data accumulate, the plan is to study how exceptions relate to specific characteristics of a particular school and curriculum and the nature of performance measures. More data will also allow further study into the methodological factors previously described that hamper accurate validity assessments.

Two further directions for MCAT validity research are in progress. The first relates MCAT scores to absolute measures of success/failure in medical school and measures indicative of academic problems. These include withdrawal/dismissal from medical school for academic reasons, delayed graduation due to repeating courses or deceleration of the academic program, and/or failure to pass Part I of the NBME examinations. These data should prove useful to those interested in establishing a threshold at which applicants may equally be expected to satisfy the academic demands of the medical school curriculum.

The second direction for research is to examine the nature and extent of relationships between MCAT scores and measures of performance in the third- and fourth-year of medical school. These are given the general label of clinical performance measures, although they represent several distinct types. Clinical knowledge measures include scores on the NBME Part II examination, NBME mini-tests, or in-house examinations used by schools in the computation of clerkship grades. These measures might reasonably be expected to correlate both with basic science knowledge measures and the MCAT. Clinical skills measures are themselves of two kinds: (*a*) those that imply a cognitive skill, for example, appropriate emphasis on pertinent facts in history-taking or the ability to integate clinical information from various sources to identify a problem, and (*b*) those that are primarily noncognitive in nature, for example, sensitivity to a patient's overall medical and personal problems and effectiveness in securing the cooperation of patient and family. Those MCAT subtests which assess skills in gathering, analyzing, and evaluating information may be expected to correlate with the former type of clinical skill measure but not the latter. However, knowing the full implications of using MCAT scores in admissions demands that relationships among all these measures be examined. This is also true with regard to the career choices made by medical students during this period. The AAMC's MCAT Interpretive Studies Program has as its goal to describe this mul-

tifaceted picture of the implications of test use.

References

1. JONES, R. F., and ADAMS, L. An Annotated Bibliography of Research on the New MCAT. *Psychol. Documents*, **13**:27, 1983.
2. LORD, F. M., and NOVICK, M. *Statistical Theories of Mental Test Scores*. Reading, Massachusetts: Addison-Wesley, 1968.
3. LINN, R., HARNISCH, D., and DUNBAR, S. Corrections for Range Restriction: An Empirical Investigation of Conditions Resulting in Conservative Correction. *J. Appl. Psychol.*, **66**:655–663, 1981.
4. GOLMON, M. E., and BERRY, C. A. Comparative Predictive Validity of the New MCAT Using Different Admissions Criteria. *J. Med. Educ.*, **56**:981–986, 1981.
5. SEDLACEK, W. E., and HUTCHINS, E. B. An Empirical Demonstration of Restriction of Range Artifacts in Validity Studies of the Medical College Admission Test. *J. Med. Educ.*, **41**:222–229, 1966.
6. LINN, R. Ability Testing: Individual Differences, Prediction and Differential Prediction. In *Ability Testing: Uses, Consequences, and Controversies*. A. Wigdor and W. Garner (Eds.). Washington, D.C.: National Academy Press, 1982, Pp. 335–388.

DATAGRAM

MCAT Scores and Student Progress in Medical School

In an article appearing in this issue of the journal, the authors summarize correlations found between Medical College Admission Test (MCAT) scores and student achievement in the first two years of medical school as measured by course grades and scores on the National Board of Medical Examiners (NBME) Part I examination. A validity question of further practical import to admissions committees is how MCAT scores relate to the probability of a student having academic problems that delay or impede his progress through medical school. As with most questions of test validity, this is best answered from local institutional experience with the MCAT scores. However, local studies are inevitably hampered by the low incidence of students who experience academic problems, presumably because of the effective use of various measures of academic achievement and student motivation in the selection process. The purpose of the authors in this datagram is to demonstrate the relationship between MCAT performance and student progress through medical school for the enrolled U.S. medical student population as a whole. While such a data aggregation may mask differences between institutions, it should provide some perspective for schools interested in establishing a minimum threshold level of performance at which a student may reasonably be expected to satisfy the demands of the medical school curriculum in the normal time period.

Method

The population studied includes all students who entered in 1978 and 1979 the 126 medical schools accredited by the Liaison Committee on Medical Education (LCME). Academic status records of each student are maintained in the Student Records System of the Association of American Medical Colleges (AAMC), beginning with the matriculation date and ending at the point of graduation or other termination. Changes in status are provided periodically by the schools to the AAMC. A majority of students entering in 1978 and 1979 were expected to graduate in either 1982 or 1983, respectively.

From information contained in the Student Records System, the authors grouped students into five categories: (*a*) graduated on time, (*b*) delayed graduation for academic reasons, (*c*) delayed graduation for nonacademic reasons, (*d*) withdrawal/dismissal for academic reasons, and (*e*) withdrawal/dismissal for nonacademic reasons.

Students were classified in group A if their actual graduation date was no later than two months subsequent to their original expected graduation date indicated at the time of matriculation. While these may include students who experienced some academic problems in medical school, these students would have managed to rectify the problems in time to graduate with their class. Students who graduated later than the expected date or who continued to be enrolled were classified in groups B or C. Those students who took a leave of absence for nonacademic reasons (for example, for health reasons) were placed in group C. The remaining students who maintained continuous enrollment were presumed to have decelerated their program or to have been made to repeat part of the curriculum because of academic difficulties. These assumptions were necessary because direct designations of the reasons for changes in expected graduation dates are not provided in the reporting system. Students who withdrew or were dismissed from medical school were classified in groups D or E, depending upon whether the school reported that the withdrawal/dismissal was for academic or nonacademic reasons. Exceptions to this were those students who subsequently reentered

TABLE 1

MCAT Chemistry Scores of 1978 and 1979 First-Year Entrants to U.S. Medical Schools by Retention and Rate of Progress

MCAT Chemistry Score	All Entrants		Graduated on Time		Delayed Graduation				Withdrawals/Dismissals			
					Academic†		Nonacademic‡		Academic		Nonacademic	
	No.	Percent	No.	Percent§	No.	Percent §	No.	Percent§	No.	Percent§	No.	Percent§
12–15	5,629	18.0	5,034	89.4	257	4.6	196	3.5	11	0.2	131	2.3
11	5,193	16.6	4,740	91.3	200	3.9	137	2.6	26	0.5	90	1.7
10	5,776	18.4	5,264	91.1	235	4.1	142	2.5	38	0.7	97	1.7
9	5,383	17.2	4,849	90.1	233	4.3	133	2.5	49	0.9	119	2.2
8	4,060	13.0	3,587	88.3	204	5.0	108	2.7	71	1.7	90	2.2
7	2,668	8.5	2,239	83.9	186	7.0	96	3.6	83	3.1	64	2.4
6	1,493	4.8	1,193	79.9	174	11.7	49	3.3	47	3.1	30	2.0
5	705	2.2	488	69.2	102	14.5	36	5.1	52	7.4	27	3.8
4	317	1.0	178	56.2	84	26.5	11	3.5	29	9.1	15	4.7
3	90	0.3	48	53.3	22	24.4	3	3.3	12	13.3	5	5.6
2	23	0.1	7	30.4	7	30.4	3	13.0	4	17.4	2	8.7
1	2	0.0	1	50.0	1	50.0	0	0.0	0	0.0	0	0.0
Total	31,339*	100.0	27,628	88.2	1,705	5.4	914	2.9	422	1.3	670	2.1

* A total of 32,354 students we. e identified as 1978 and 1979 first-year entrants. This table excludes 551 students enrolled in B.A./M.D., M.D./M.A., M.D./Ph.D., and M.D./J.D. programs, 51 deceased students, and 413 students without new MCAT scores.

† Includes repeaters, students in decelerated programs, and academic withdrawals/dismissals who reentered medical school.

‡ Includes those who took leaves of absence or delayed progress in transferring.

§ Percentages based on the total number of students at each particular score level.

TABLE 2

MCAT Skills Analysis: Reading Scores of 1978 and 1979 First-Year Entrants to U.S. Medical Schools by Retention and Rate of Progress

MCAT Reading Scores	All Entrants		Graduated on Time		Delayed Graduation				Withdrawals/Dismissals			
					Academic†		Nonacademic‡		Academic		Nonacademic	
	No.	Percent	No.	Percent§	No.	Percent§	No.	Percent§	No.	Percent§	No.	Percent§
12–15	2,651	8.5	2,378	89.7	123	4.6	84	3.2	9	0.3	57	2.2
11	4,130	13.2	3,705	89.7	181	4.4	124	3.0	17	0.4	103	2.5
10	8,452	27.0	7,643	90.4	348	4.1	231	2.7	68	0.8	162	1.9
9	5,921	18.9	5,310	89.7	271	4.6	167	2.8	56	0.9	117	2.0
8	5,384	17.2	4,768	88.6	286	5.3	137	2.5	81	1.5	112	2.1
7	2,749	8.8	2,332	84.8	201	7.3	87	3.2	61	2.2	68	2.5
6	868	2.8	694	80.0	77	8.9	30	3.5	37	4.3	30	3.5
5	691	2.2	492	71.2	110	15.9	34	4.9	43	6.2	12	1.7
4	249	0.8	162	65.1	47	18.9	15	6.0	22	8.8	3	1.2
3	144	0.5	86	59.7	34	23.6	4	2.8	18	12.5	2	1.4
2	53	0.2	30	56.6	14	26.4	1	1.9	5	9.4	3	5.7
1	47	0.1	28	59.6	13	27.7	0	0.0	5	10.6	1	2.1
Total	31,339 *	100.0	27,628	88.2	1,705	5.4	914	2.9	422	1.3	670	2.1

* A total of 32,354 students were identified as 1978 and 1979 first-year entrants. This table excludes 551 students enrolled in B.A./M.D., M.D./M.A., M.D./Ph.D., and M.D./J.D. programs, 51 students who are deceased, and 413 students without new MCAT scores.

† Includes repeaters, students in decelerated programs, and academic withdrawals/dismissals who reentered medical school.

‡ Includes those who took leaves of absence or delayed progress in transferring.

§ Percentages based on the total number of students at each particular score level.

medical school; they were classified in groups B or C, depending upon the reasons for the withdrawal/dismissal. In order to classify students into these five categories, students enrolled in B.A./M.D., M.D./M.A., M.D./J.D., or M.D./Ph.D. programs and others whose original expected graduation date was later than November 1983 were dropped from the analyses, since graduation data would not have been available at the time the analyses were performed. In addition, students who died during the study period and students without new MCAT scores were not included.

Findings

Tables 1 and 2 show for two selected areas of assessment, Chemistry and Skills Analysis: Reading, the number and proportion of students at each MCAT score level who were classified into each of these five groups (tables based on scores for other MCAT areas of assessment are available from the authors). The results indicate that 88.2 percent of the 1978 and 1979 first-year entrants graduated on time. A total of 8.3 percent had their graduation delayed, 5.4 percent for academic reasons and 2.9 percent for nonacademic reasons. Of the 3.4 percent of the students who either withdrew or were dismissed, 1.3 percent were for academic reasons while 2.1 percent were for nonacademic reasons. The data reveal a positive relationship between performance on the MCAT and graduating on time and an inverse relationship between MCAT performance and both delayed graduation for academic reasons and withdrawal/dismissal for academic reasons. The proportion of students withdrawing or delaying progress for nonacademic reasons tended to be fairly uniform across the range of MCAT scores. The probability of experiencing academic difficulties showed little variation for those students achieving scores between 8 and 15 on the MCAT Chemistry and Skills Analysis: Reading subtests. However, for those scoring below 8 on either subtest, the probability of encoun-

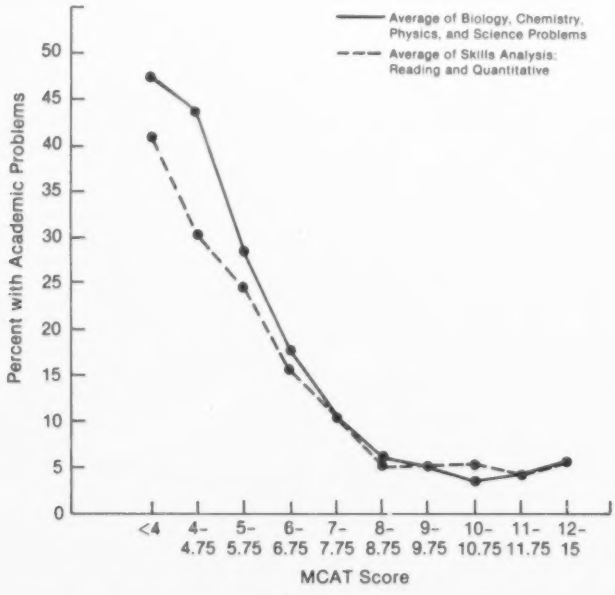

FIGURE 1

Percentage of students having academic problems in U.S. medical schools by Medical College Admission Test (MCAT) scores, 1978 and 1979 entering classes.

tering academic problems tended to increase systematically as MCAT scores decrease.

Figure 1 illustrates further the relationship between MCAT performance and the likelihood of experiencing academic problems by comparing scores for the Science Knowledge and Science Problems subtests with those for the Skills Analysis: Reading and Skills Analysis: Quantitative subtests. For simplicity of presentation, students whose academic problems caused a delay of graduation or withdrawal/dismissal were grouped together, and MCAT scores were averaged within the science and skills areas. For both averages, the likelihood of a student experiencing academic problems increased consistently as scores declined below 8 on either measure. The increase tended to be slightly more pronounced as the Science Knowledge and Science Problems subtests scores decline. Since academic problems tend to surface in the first year of medical school, this finding is not surprising. For both the old and the new MCAT, the strongest correlations with first-year performance have been found for MCAT scores in the science areas of assessment.

Even at the very lowest levels of MCAT performance, approximately half of the small numbers of students who were accepted were successful in graduating from medical school on time. This appears to reflect the effectiveness of admissions committees in identifying other factors that predict student success in addition to MCAT scores. The graduation rate also argues against an absolute and rigid use of MCAT scores in admissions decisions but for its use as contributing information to a complete applicant profile.

ROBERT F. JONES, PH.D., *research associate, and* SUZANNE VANYUR, *research assistant, Division of Educational Measurement and Research, Association of American Medical Colleges, Washington, D.C.*